UDC 004.852 **doi: 10.32620/reks.2024.3.14**

Zoia DUDAR, Andrii KOZYRIEV

National University of Radio Electronics, Kharkiv, Ukraine

FINITE PREDICATE-DRIVEN LOGIC NETWORKS METHOD FOR ENHANCED EDUCATION DATA ANALYSIS

*The **subject matter** of the study is intelligent data analysis in the field of academic information. The **goal** of the study is to create a mathematical model for analyzing students' academic information using the predicate-driven logic networks method, which allows taking into account both logical dependencies and probabilistic transitions between states. To achieve this goal, the following **tasks** were defined: analysis of the theoretical foundations of the logic networks method and predicate logic, integration of these approaches into a single mathematical model, development of approaches for its application in academic data analysis problems. The research used the **methods** of mathematical modeling, complex logical analysis, and method for constructing logic networks. The following **results** were obtained: a theoretical model was developed that integrates the principles of logic networks and predicate logic for analyzing student academic performance; the model accounts for both probabilistic transitions between states and logical dependencies among student parameters; the mathematical model also incorporates logical rules to enhance the accuracy of logical analysis within the academic context. The model was tested on a dataset of student performance, demonstrating its effectiveness in accurately predicting academic outcomes and confirming the validity of the integrated approach. **Conclusions**. The scientific novelty of the results obtained is as follows: 1) a theoretical model for analyzing student academic data was developed by integrating logic networks and predicate logic, allowing for the simultaneous consideration of probabilistic transitions and logical dependencies among student parameters; 2) the approach enhances the analysis process by incorporating logical rules into the probabilistic framework, providing a more nuanced and accurate tool for analyzing academic data; 3) this combined model offers a novel method for addressing complex logical analysis tasks in educational settings, paving the way for further research and practical applications in intelligent data analysis. The successful testing of the model on actual student data further underscores its potential as a powerful tool in educational data analysis.*

Keywords: *Intelligent data analysis; academic data; logic networks; algebra of finite predicates; predicate logic.*

Introduction

Intelligent data analysis has become an important component of many areas of science and technology, especially in the context of big data. In the educational field, digitalization has led to a significant increase in data related to students' academic performance, their activity, attendance, participation in scientific and extracurricular activities. This requires the use of advanced analytical methods that could effectively process and interpret this information, providing a deeper understanding of educational processes and student success.

Many existing approaches, such as statistical analysis, neural networks, and machine learning methods, have proven their effectiveness in solving standard data analysis and forecasting problems. However, these methods are often not flexible enough to take into account both probabilistic and logical dependencies between various parameters, such as academic performance, attendance, and other student characteristics.

In this regard, there is a need to develop new models that could not only process large amounts of data, but also take into account complex logical and probabilistic relationships between various states and characteristics of students in the process of their learning.

This work is aimed at creating and studying a new mathematical model based on the method of logical networks controlled by predicates. This method allows taking into account both logical dependencies between student parameters and probabilistic transitions between states, which makes it promising for more accurate data analysis and forecasting of academic performance.

1. Motivation

Modern educational systems, especially in the context of virtual universities and digital learning, are faced with the need to effectively analyze large volumes

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

of academic data. Without proper analysis, such data remains unstructured and underutilized to make informed decisions on improving the educational process [1, 2]. Virtual learning platforms and digital learning systems generate a huge amount of information, including data on student performance, their participation in educational programs, attendance, and many other parameters.

In the era of digitalization of education, traditional statistical analysis methods are often insufficient to identify complex relationships and trends in academic data. Simply accumulating information, such as academic records of millions of students, does not allow for a deep understanding of the dynamics of their learning and engagement. In educational analytics, it is important not only to collect data, but also to build models that can interpret and analyze this information in order to create recommendations based on it for improving the educational process [3].

Intelligent analysis of academic data using advanced mathematical methods allows you to identify hidden patterns and trends that remain unnoticed by traditional approaches. This is especially important in the context of digital learning, where data systematization is necessary for their deep analysis and effective use [4].

In this context, the method finite predicate-driven logic networks offers a new approach to solving the problems of analyzing students' academic performance. This method allows analyzing and analyzing data, taking into account both logical dependencies between students' parameters and probabilistic transitions between different states of their academic activity.

2. State of the Art

Intelligent data analysis employs mathematical techniques to uncover patterns and trends that traditional methods often fail to identify due to the complexity of the relationships or the vastness of the data involved. By defining and integrating these patterns into a cohesive model, researchers can achieve a deeper understanding and more effective utilization of existing data. Such analysis methods are particularly crucial when dealing with logical analysis, where the process of analyzing objects requires careful selection of characteristic values to determine the object's membership in a specific class.

As an example, the study [5] performs classification of binary feature vectors. The proposed methodology finds application in a wide range of data analysis problems in various fields. The classification process is based on the analysis of logical-dynamic systems, considering certain initial states.

The research detailed in [6] developed specialized correction functions for methods used in logical object identification, presenting an innovative logical analysis algorithm that integrates these functions to refine the precision of object analysis based on logical principles.

Another significant contribution to the field is presented in [7], which explores the application of Boolean data analysis of hyperspectral images. This method uses the binary simplicity of Boolean algebra to manage the intricate data structures inherent in hyperspectral imaging, facilitating more effective data interpretation.

The study in [8] delves into the combinatorial aspects of inferential analysis. Here, the authors applied their theoretical insights to the practical challenge of classifying proteomic expressions linked to Alzheimer's disease, demonstrating the potential of combinatorial logic in addressing complex biological data analysis.

Furthering the application of logical methods, the work in [9] describes logical algorithmic techniques for constructing decision trees, a fundamental tool in various decision-making processes across different industries, from healthcare to finance.

In the realm of text analysis, [10] proposes a logic-based quantitive analysis method specifically designed for text recognition. This approach adapts logical principles to the challenges posed by textual data, illustrating the adaptability of logical methods to diverse types of data.

The study in [11] introduces a logic approach to machine learning, including the development of a specialized logic classifier. This approach aims to integrate logical reasoning into machine learning frameworks, enhancing model interpretability and accuracy.

The concept of combining multiple weak classifiers to design an effective overall classifier is explored in [12]. This research discusses ensemble methods, which leverage the collective strength of several simple models to achieve superior data analysis performance.

An extensive integration of various methods, including logical analysis, was empirically tested in a study referenced in [13]. This integration underscores the potential benefits of hybrid approaches in solving more complex data analysis tasks.

Text message data analysis based on logical inference is the focus of [14], highlighting how logical methods can be effectively utilized for the real-time analysis of dynamic text data.

In the field of radar technology, [15] examines finite predicate networks, demonstrating the application of logical networks in environments where rapid and accurate data processing is critical.

Innovative techniques in logical analysis and recognition are presented in [16], where new approaches are explored to refine the methodologies used in logical data analysis across various applications.

The study in [17] explores the subsystems of systems of Boolean equations, providing deeper insights into the nuances of Boolean systems that are crucial for understanding and developing complex logical models.

A method for the distributive solution of logical equations is introduced in [18], aiming to enhance the efficiency and feasibility of solving logical equations, which are foundational to many computational logic applications.

Special attention is given to specific types of logical functions, particularly symmetric functions, in [19]. These functions are actively used in data analysis tasks due to their unique properties that simplify computational processes.

Entropy issues within Boolean networks are covered in [20], discussing the theoretical aspects of entropy and its implications for the design and functionality of logical networks.

Studies in [21, 22] further investigate logic-predicate networks, enhancing our understanding of these networks in computational settings.

Lastly, [23] discusses the application of predicates in classifying access issues within the Internet of Things (IoT) domain, illustrating the broad applicability of logical approaches to new and emerging technologies.

These studies collectively underscore the ongoing evolution and integration of logical and algorithmic methods in addressing diverse challenges across various scientific and technological disciplines. They highlight the convergence of classical logic theories with modern computational technologies to tackle contemporary data analysis and classification challenges.

In the realm of logical analysis and algorithmic methods, extensive research has significantly advanced the application of logic in data analysis. A wide array of studies has explored various dimensions of logical frameworks, each contributing to the collective understanding and enhancing the practical implementation of these methodologies in addressing complex real-world challenges. In this context, the application of predicate-driven logic netwroks offers a powerful extension to the existing state of the art, providing a robust mechanism for modeling dynamic transitions in academic performance. By integrating multiple factors such as attendance, grades, scientific activity, and extracurricular involvement, this method allows for a more nuanced analysis of student success over time.

3. Objective and Approach

The main goal of our research is to create and validate a mathematical model of predicate-driven logical networks (PDLN) that effectively analyzes and predicts results based on students' academic data. We aim to develop a tool that can accurately predict academic outcomes, identify key factors that influence academic performance, and propose data-driven solutions to improve educational processes. This model is designed for in-depth analysis of multiple, interrelated academic parameters such as academic performance, participation in academic and social activities, and other relevant indicators.

The study is based on an integrated approach that combines theoretical research and practical implementation to ensure a comprehensive analysis and effective application of the proposed model:

1. Formation of the theoretical framework: A detailed analysis and synthesis of the literature on predicate logic and logical networks is conducted. This includes studying current research in this area to identify existing problems and possible solutions through new technological approaches.
2. Development of a mathematical model: Based on the preliminary analysis, a mathematical model is created that integrates the principles of logical networks and predicate logic. This model is designed to handle complex probabilistic state transitions and logical dependencies between different academic parameters.
3. Practical application and testing of the model: The model is tested on real student data, which includes data collection, data cleaning, and preparation for analysis. The model is then applied to analyze and predict academic outcomes, which allows us to evaluate its accuracy and applicability in real-world settings.
4. Result analysis and performance evaluation: The results of the model testing are analyzed to evaluate its effectiveness in different educational contexts. An important aspect is to evaluate the model's ability to identify complex patterns and provide interpretable findings that can be used to improve the educational process.

The research methodology combines deep theoretical analysis with practical application, which makes it possible not only to create a reliable and functional model, but also to ensure its relevance and effectiveness for the modern educational process. This approach allows not only to solve existing data analysis problems, but also to suggest ways to overcome them using the latest technologies.

The article is organized as follows:

1. Introduction, which substantiates the relevance of the study.
2. State of the Art, which provides an overview of existing methods and approaches in the analysis of academic data.
3. Objective and Approach, which describes the goals and approaches of the study.
4. Materials and Methods, which describes the methods and materials used in the study, including the

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

basics of predicate logic, its application to data analysis, and integration with logical networks.

5. Case Study, which demonstrates the application of the model in a practical example and analyzes the results obtained.

6. Results and Discussion, which discusses the results, compares them with other methods, and identifies potential improvements.

7. Conclusion, which summarizes the results and describes directions for future research.

4. Materials and Methods

4.1. Predicate Logic in the Context of Data Analysis

An extension of Boolean algebraic equations is finite predicate algebraic equations, which provide the ability to work with attribute variables defined on finite sets. The use of these equations to create logical conclusions in knowledge base systems expands the possibilities of logical identification of objects. Data and their relationships can be represented in many ways, but first consider the basics of Boolean algebra and the challenges associated with resolving them.

Certain mathematical structures that satisfy the axioms of Boolean algebra play a key role in information systems, including the algebra of logic and set algebra. The dependencies between elements of Boolean algebra can often be described by one or more equations involving both unknown and known elements. It is important to determine the conditions under which these equations have solutions and when these solutions are unique, and to find method for determining them.

When the conditions for uniqueness of a solution are not met, finding all possible solutions to an equation or system of equations can be difficult due to their large number. However, there is interest in studying these solutions, choosing the optimal ones for certain problems, and analyzing their structure, like the analysis in the theory of differential equations. As will be shown later, there is a method for determining all solutions to Boolean equations without the need to brute force them sequentially, describing them using formulas from which all solutions can be obtained.

Variables such as $y_1, y_2, \ldots, y_m$, reflect certain characteristics of objects. This may include, as an option, a structured presentation of criteria for assessing the situation, for example, according to the norms of quality assessment or categorization of the subject according to levels of significance. In contrast to Boolean variables, which are limited to two states, predicate variables allow for a wide range of variability due to their unique, discrete ranges of values. Considering the case when discrete variables $x_1, x_2, \ldots, x_n$ are characteristics by which we can estimate potential values of object characteristics [24], these characteristics and signs can be interconnected through complex logical relationships, which allows you to formulate them through predicate equations:

$$P(y_1, y_2, \ldots, y_m;\ x_1, x_2, \ldots, x_n) = 1 \qquad (1)$$

The analysis of this object consists in determining its properties (or lack thereof) based on the presented predicate equation and the obtained experimental data on the characteristics $x_1, x_2, \ldots, x_n$. Specifically, through the analysis of each basic conjunction, it is possible to establish the attractiveness of an object to a certain class. Based on the predetermined dependence (1) and data on the characteristics $x_1, x_2, \ldots, x_n$, it is possible to clarify the analysis of the object. These characteristics are organized in the form of a matrix for ease of analysis. Based on the experimentally obtained data on the characteristics $x_1, x_2, \ldots, x_n$ that characterize the analyzed object, a predicate equation is formulated that reflects the relationships between these characteristics [24, 25].

The process of analyzing objects can be formally presented through the solution of a given predicate equation to determine the unknown predicate f:

$$g(x_1, x_2, \ldots, x_n) \rightarrow f(y_1, y_2, \ldots, y_m) \qquad (2)$$

Having solved this functional equation, it is possible to determine the values of the features $x_1, x_2, \ldots, x_n$ that characterize the objects $y_1, y_2, \ldots, y_m$.

Predicate logic provides much more opportunities for describing the characteristics of objects compared to Boolean logic, which is limited to only two states: true or false. In predicate logic, each variable can take values from a whole set, which allows for a significant increase in the complexity and expansion of modeled dependencies.

An example of a dependency between variables can be described as follows: variable y can take values from a certain set, say $\{a_1, a_2, a_3\}$, and variable x_1 can take values from another set, say $\{b_1, b_2, b_3, b_4, b_5, b_6\}$. This creates conditions for defining a dependency between these variables. For example, if y takes the value a_1, then this automatically means that variable x_1 can only take the values b_1 or b_2. In other words, there is a logical relationship that limits the possible values of the variables based on their mutual dependencies.

This relationship can also be interpreted through negation. If it is known that neither b_1 nor b_2 can be true for x_1, then this can lead to the conclusion that y cannot be equal to a_1. From this we can conclude that under such conditions the variable x_1 must take one of the values from b_3 to b_6, and the variable y will accordingly go into one of the states a_2 or a_3. Thus, predicate logic allows us to describe complex dependencies between a set of variables at the level of logical expressions.

The process of feature selection (or feature selection) can be thought of as a method that allows reducing the number of variables used in the analysis. This is achieved using so-called substitution operators. A substitution operator works as follows: if a predicate P depends on a set of variables, for example (x, y, z), then the substitution operator $a(P)$ allows one of the variables to be replaced by a fixed value, while the predicate retains its truth value for the remaining variables. For example, if the predicate P depends on the variables (x, y, z), and we apply the substitution operator to the variable x, replacing it with a, this transforms the predicate as follows: $P(x, y, z)$ becomes $P(a, y, z)$.

Such an operator can be restrictive if, for any combination of variables, the truth of the predicate after the substitution always leads to the truth of the predicate before the substitution. In other words, the restrictive operator ensures that the substitution does not violate the logical relationships between the variables.

Substitution operators can also be distributive. This means that substitution can not only preserve logical dependencies, but also weaken them by changing the connections between variables. For example, if the predicate P describes a dependency between several variables, the substitution operator can change the structure of the predicate in such a way that the logical connections between the variables are weakened, which leads to a simplification of the predicate.

For example, the predicate $P(x, y, z)$ can be written as a combination of several logical expressions that depend on the variables x, y, and z. If the substitution operator is applied to the variable x, then we can rewrite the predicate, simplifying it so that the dependencies on x disappear, but the connections between the variables y and z remain.

Restrictive substitution operators strengthen the logical connection between variables, preserving all dependencies between them, while distributive operators weaken the connection, allowing more flexibility in constructing the predicate. For example, if $P(x, y, z)$ expresses a dependency between three variables, and we apply the distributive operator to x, this may result in the dependence on x becoming less rigid, and the other variables receiving more freedom in their values.

Thus, predicate logic in combination with substitution and distribution operators provides a powerful tool for working with large and complex data systems, allowing you to manage multiple interdependent parameters and simplify their analysis.

4.2. Integrating Logic Networks with Predicate Logic

Logic networks model sequences of states where the transition to the next state depends only on the current state (Markov property) [26, 27]. In the context of student's educational data analysis task, states may represent current performance, behavior, or other parameters.

The logic network is described by the transition probabilities between states:

$$P(X_{t+1} \mid X_t) = P(X_{t+1} \mid X_t, X_{t-1}, ..., X_p), \qquad (3)$$

where X_t – state of the system at time t.

The idea is to use predicates to determine the probabilities of transitions in a logic network. This means that transitions between states will depend not only on the current state, but also on the fulfillment of certain logical conditions.

States of a logic network can be defined as sets of predicate values. The state S_t can be defined by satisfying a number of predicates:

$$S_t = \{P_1(x), P_2(y), ...\}. \qquad (4)$$

Transitions between states will be determined by predicates. For example, the transition from the state S_t to S_{t+1} will depend on the fulfillment of the predicates:

$$P(S_{t+1} \mid S_t) = P(X_{t+1} \mid X_t) \cdot P(P_1(x), P_2(y), ...), \qquad (5)$$

where $P(P_1(x), P_2(y), ...)$ is a probability of fulfilling logical conditions.

Transition probabilities may change depending on the execution of predicates. For example, if the predicate $P_p(x)$ corresponds to the fact that the student regularly attends classes, then the probability of moving to the state of "successful passing of the exam" will be higher.

$$P(S_{t+1} \mid S_t, P_p(x) = \text{true}) > P(S_{t+1} \mid S_t, P_p(x) = \text{false}) \qquad (6)$$

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The model can be described as a set of logical states with predicates that determine transitions:

$$P(S_{t+1} \mid S_t) = \sum_{n=1}^{n} w_n P(P_n(x)) \tag{7}$$

where w_n is a weighting function that considers the importance of each predicate.

5. Case Study

We can apply the PDLN model to analyze students' academic data to evaluate the impact of different parameters on academic performance.

We analyzed data of 150 students, designated as $S_1, S_2, ..., S_{150}$. For each student S_n the following parameters are known:

- N – number of students, $N = 150$;
- M – number of semesters, $M = 8$;
- S'_n – student n, where $n = 1, 2, ..., N$;
- $G_{n,m}$ – student's S_n midterm grade for the semester m;
- $A_{n,m}$ – student's S_n attendance for the semester m measured in hours;
- $P_{n,m}$ – number of student's S_n publications;
- $C_{n,m}$ – student's S_n participation in competitions and Olympiads during the semester m, where $C_{n,m} \in \{\text{Low,Mid,High}\}$.

A student's state in a semester is defined as a combination of all four parameters:

$$S_{n,m} = (G_{n,m}, A_{n,m}, P_{n,m}, C_{n,m}) \tag{8}$$

where $S_{n,m}$ is a tuple, that describes students' S'_n state in semester m.

For example, $S_{n,m} = (\text{Mid,High,Low,Mid})$ describes a student with a state of "Mid" for midterm grade, "High" for attendance, "Low" for number of publications, and "Mid" for participation in competitions in the semester.

Transition probability T_{kp} is calculated as:

$$T_{kp} = P(S_{n,m+1} = S_p \mid S_{n,m} = S_k) \tag{9}$$

where, $P(S_{n,m+1} = S_p \mid S_{n,m} = S_k)$ the probability that a student with state S_k in semester m will move to state S_p in semester $m+1$.

To model the change in the student's state during their learning, a PDLN model is used. Let:

- $\mathbf{S}$ – the set of all possible states of a student, consisting of combinations of all four parameters;
- $\mathbf{T}$ – transition matrix of size $|\mathbf{S}| \times |\mathbf{S}|$, where T_{kp} denotes the probability of transition from the state S_k into state S_p.

The final state of student $S_{n,M}$ after all semesters is defined as:

$$S_{n,M} = \arg\max_p \left(\prod_{m=1}^{M-1} T_{S_{n,m},p} \right) \tag{10}$$

where $\prod_{m=1}^{M-1} T_{S_{n,m},p}$ is the product of the probabilities of transition between states for a student S_n over all semesters.

To implement the proposed mathematical model, a program was written in Python, which allowed for detailed analysis and visualization of data.

To simplify the calculations and reduce the number of unique states, the model was optimized by using the arithmetic mean to encode student states across four parameters: GPA, attendance, publication activity, and competition participation.

The state now can be defined as arithmetic mean of all four parameters:

$$S_{n,m} = \frac{G_{n,m} + A_{n,m} + P_{n,m} + C_{n,m}}{4} \tag{11}$$

where is a numeric value, that describes students' S'_n state in semester m.

The definition of parameters is similar to the previous description, but with numerical values of parameters: $G_{n,m} \in \{1, 2, 3\}$, $A_{n,m} \in \{1, 2, 3\}$, $P_{n,m} \in \{1, 2, 3\}$, $C_{n,m} \in \{1, 2, 3\}$, where 1 = Low, 2 = Mid, 3 = High.

The model is similar to the previous case (9), but now the states S_p and S_k are numeric values.

Using the arithmetic mean of the parameters significantly reduces the number of unique states, limiting them from 81 to single non-discrete value. This simplifies the calculations and reduces the size of the transition matrix, making the model more efficient in terms of computational resources.

Logic network, presented on Figure 1 reflecting the probabilities of transitions between different states of students during the course of study for 8 semesters. Each node on the graph represents the arithmetic mean of the students' states, ranging from 1.0 to 3.0, where 1.0 corresponds to low values of all parameters (grade point average, attendance, publications, participation in competitions), and 3.0 corresponds to high values of these parameters. Arrows between nodes display possible transitions of students from one state to another, and the probability of such a transition is indicated next to each arrow.

The chain demonstrates that transitions between states with similar values are most probable, reflecting the stability of students in certain categories. For example, students in a state with a low midterm grade (1.0) are likely to remain in this state or move to a slightly improved state (1.25 or 1.5). Conversely, transitions to higher states (e.g., from 1.5 to 3.0) are less probable and require significant changes in student behavior across multiple parameters at once.

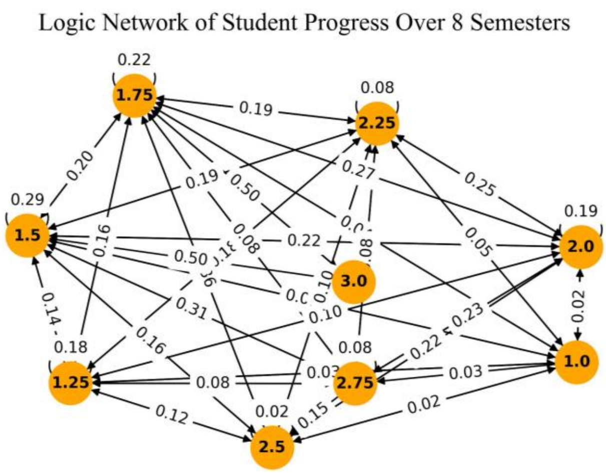

Fig. 1. Logic network transitions probabilities

The distribution graph of the final states (Figure 2) shows that most students are concentrated in the range from 1.5 to 2.0 by the average state score. This indicates that most students have states close to "Mid" (2) by the set of parameters (average score, attendance, publications, participation in competitions).

Few students achieved values above 2.5, indicating the difficulty of achieving a high state, possibly due to the difficulty of achieving high values on multiple dimensions at once.

The average transition probability plot (Figure 3) shows that most students have an average transition probability in the range of 0.14...0.24.

This indicates that transitions between states are not easy and the highest probability remains stable. The different colors of the dots on the plot represent the final state of the students. The higher the value of the final state, the greater the probability of successful completion of the education.

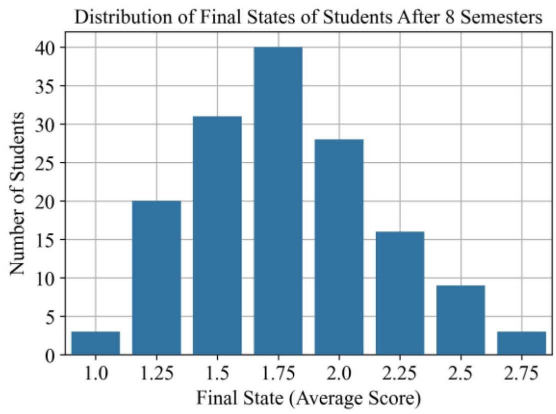

Fig. 2. Distribution of final states after 8 semesters

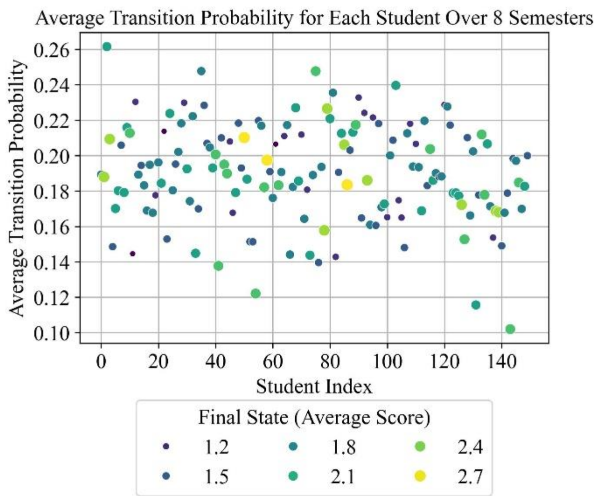

Fig. 3. Average transition probability for each student

The feature importance plot (Figure 4) for successful completion of the education was calculated using the multiple linear regression method. In this model, the target variable is the final state of the student after 8 semesters, which was coded by numerical values based on the arithmetic mean of four parameters: midterm grade, attendance, number of publications, and participation in competitions.

Linear regression was used to estimate the influence of each of the four parameters on the students' final state. The regression model takes the following form:

$$y_n = \beta_0 + \beta_1 \cdot A_{n,8} + \beta_2 \cdot G_{n,8} + \beta_3 \cdot P_{n,8} + \beta_4 \cdot C_{n,8} + \varepsilon_n \quad (12)$$

where: $A_{n,8}, G_{n,8}, P_{n,8}, C_{n,8}$ are the attendance, grade point average, number of publications and participation in competitions values of student in the eighth semester;

$\beta_1, \beta_2, \beta_3, \beta_4$ are regression coefficients estimating the contribution of each parameter;

ε_n – the model error or residual in a linear regression.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

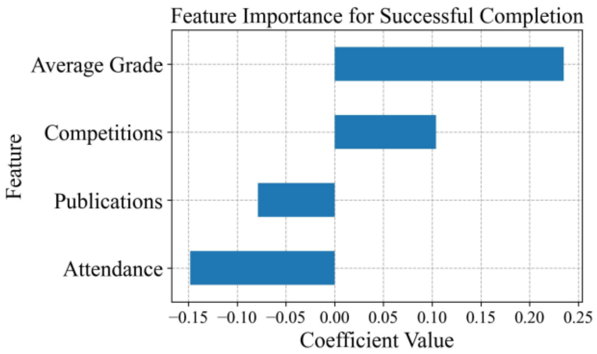

Fig. 4. Feature importance for successful completion

The obtained coefficients reflect the degree of influence of each parameter on the final state of the student. The higher the coefficient, the greater the contribution of the corresponding parameter to the successful completion of training. Participation in Competitions and Average Grade have the greatest positive impact on successful completion of studies. Attendance and Publications have a negative impact on successful completion of studies. This may mean that high levels of scientific activity and attendance do not always correlate with a high final state in the model.

Summary. The final states of most students are concentrated in the average grade range (1.5...2.0), indicating a balanced level of academic performance across all parameters.

The average transition probability indicates the stability of the states of most students, where significant changes (e.g., a jump from Low to High) are rare.

The analysis of the importance of features emphasizes that participation in competitions and academic grades are most important for achieving a successful outcome, while high attendance and publication rates do not always lead to successful completion of studies.

Results and Discussion

The results of the study showed that the application of the predicate-driven logical networks (PDLN) method to the analysis of academic data allows achieving high accuracy in analyzing and predicting students' academic success.

During the experiments, the state of 150 students was analyzed by four key parameters: GPA, attendance, publication activity, and participation in competitions. The probabilities of transitions between student states demonstrate that transitions between states with close parameter values are most likely. For example, students with a low GPA (around 1.0) more often remain in this category or move to states with slightly improved indicators (1.25 or 1.5). Statistical significance and distribution of states. The distribution of the final states of students after 8 semesters showed that most students are concentrated in the range from 1.5 to 2.0 in GPA. This indicates a general trend towards average academic performance. Only a small proportion of students (less than 10%) were able to achieve values above 2.5, which highlights the difficulty of achieving high scores on all parameters at once. Average transition probability and feature significance. As shown in Figure 3, the average transition probability between states for most students ranges from 0.103 to 0.234. This indicates that the states of most students are stable, with significant changes, such as transitions from a low to a high state, occurring rarely. Feature significance analysis showed GPA (coefficient 0.234) and participation in competitions (coefficient 0.103) have the greatest impact on successful completion of training. Interestingly, such parameters as attendance and publication activity were negatively correlated with successful completion of training (coefficients -0.148 and -0.07, respectively). This may indicate that high activity in research and attendance do not always lead to high final results, possibly due to the redistribution of students' time and resources.

Unlike traditional methods such as linear regression or neural networks, the proposed PDLN model provides a more detailed analysis by integrating both probabilistic transitions and logical dependencies. This allows for a better consideration of the dynamics of changes in students' academic performance, which is confirmed by the experimental results.

Conclusion

The application of advanced data analysis method, such as predicate-driven logic networks, has significantly expanded the capabilities of academic performance evaluation, particularly in understanding and predicting student success over multiple semesters. This method allows for a more refined analysis by integrating complex parameters, including attendance, grades, scientific activity, and extracurricular involvement, into a unified model that tracks the progression of students through their academic journey.

One of the key strengths of this approach lies in its ability to model the dynamic transitions between different states of student performance, offering insights that go beyond traditional static analysis. By simulating these transitions, researchers can uncover hidden patterns and relationships that influence academic outcomes, thereby providing a more comprehensive understanding of the factors that contribute to student success.

The optimization of the model through the use of arithmetic averaging of states not only enhances computational efficiency but also preserves the analytical depth necessary for meaningful interpretation of the results. This balance between complexity and practicality ensures that the model remains robust and applicable to a wide range of educational scenarios.

Moreover, the use of this methodology in analyzing academic data opens up new avenues for educational research, offering the potential to improve student support systems, refine academic advising practices, and ultimately enhance educational outcomes. As data-driven decision-making becomes increasingly critical in educational institutions, the method of predicate-driven logic networks provide the tools needed to transform raw data into actionable insights.

This approach also highlights the importance of integrating multiple aspects of student life into performance evaluations, acknowledging that academic success is multifaceted and influenced by a combination of academic, social, and extracurricular factors. By leveraging these comprehensive analytical techniques, educators and administrators can develop more targeted interventions and support mechanisms, thereby fostering an environment that is conducive to the holistic development of students.

Future research development in this area could explore the extension of predicate-driven logic networks to include more sophisticated variables, such as psychological and socio-economic factors, which also play a crucial role in student performance. For instance, incorporating metrics like stress levels, mental health indicators, and family background could provide a more holistic view of the factors influencing academic success. These additional variables could allow the model to capture more subtle interactions between a student's emotional well-being and their academic outcomes, paving the way for a deeper understanding of how non-academic factors shape performance. For example, tracking fluctuations in stress levels throughout the semester could reveal correlations between high-stress periods and drops in academic performance, allowing educators to intervene proactively. Integrating machine learning algorithms with predicate-driven models could enhance predictive accuracy, allowing for real-time adjustments to educational strategies based on ongoing student performance. Techniques such as reinforcement learning or deep learning could be leveraged to optimize the feedback loop, where educational interventions are continuously refined based on the model’s predictions. For example, a machine learning-enhanced model could predict when a student is at risk of academic failure and suggest personalized learning resources or tutoring before the student reaches a critical point. By training the model on large datasets, such systems could dynamically adjust the level of support provided to each student, personalizing learning pathways in a way that is currently beyond the capabilities of static models.

Applying this methodology across different educational contexts, such as online learning environments and diverse cultural settings, could provide deeper insights into the universality and adaptability of the model, thus broadening its impact and utility in the field of education. For example, the model could be tested in virtual classrooms, where student engagement metrics—such as participation in online forums or completion of digital assignments—could be factored into the analysis. Similarly, deploying the model in schools across different countries could highlight how cultural differences impact educational outcomes and whether certain parameters, like socio-economic status, play a more or less prominent role in different regions. This cross-context analysis could lead to a more globally adaptable model that supports a wide range of educational systems, from urban to rural and from high-income to low-income settings.

Future research could investigate the integration of additional data sources, such as sensor-based data from wearable technology, which could measure a student’s physical activity or sleep patterns, further enriching the model's understanding of factors that influence academic performance. For instance, real-time tracking of sleep cycles could help identify students who are sleep-deprived and may benefit from adjustments in workload or additional mental health support. Exploring collaborations between educators, data scientists, and psychologists could also lead to the development of hybrid models that merge traditional educational theories with advanced computational techniques. Such interdisciplinary research would not only strengthen the theoretical foundation of predicate-driven logic networks but also improve their practical implementation in diverse educational scenarios.

In summary, the future of predicate-driven logic networks lies in expanding their complexity and reach, integrating new variables, leveraging machine learning, and testing the model across various educational and cultural settings. This evolution will undoubtedly lead to more precise, personalized, and adaptable tools for enhancing student outcomes in the digital age.

Contributions of authors: conceptualization, methodology, formulation of tasks – **Zoia Dudar;** analysis – **Andrii Kozyriev;** development of model, software, verification – **Andrii Kozyriev;** analysis of results, visualization – **Andrii Kozyriev;** writing – original draft preparation visualization – **Andrii Kozyriev;** writing – review and editing – **Zoia Dudar.**

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, author ship or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted without financial support.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Data Availability

The manuscript has no associated data

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

All authors have read and agreed to the published version of this manuscript.

References

1. West, S. W., Clubb, J., Blake, T. A., Fern, J., Bowles, H. and Dalen-Lorentsen, T. Big data. Big potential. Big problems? *BMJ Open Sport & Exercise Medicine*, 2024, vol. 10, iss. 2, article no. e001994. DOI: 10.1136/bmjsem-2024-001994. (accessed 1 August 2024)
2. Jadhav, S. S. Big Data Analytics. *International Journal for Research in Applied Science and Engineering Technology*, 2024, vol. 12, iss. 4, pp. 843-847. DOI: 10.22214/ijraset.2024.59806.
3. Ravi, V. & Cherukuri, A. K. *Handbook of Big Data Analytics: Volume 1: Methodologies*, 2021. 391 p. DOI: 10.1049/PBPC037F.
4. Rayala, V. & Kalli, S. R. Big data clustering using Improvised Fuzzy C-Means clustering. *Revue d'Intelligence Artificielle*, 2021, vol. 34, iss. 6, pp. 701-708. DOI: 10.18280/RIA.340604.
5. Koenig, M. E. D. *What is KM? Knowledge management explained*. KMWorld 2024 Is Nov. 18-21 in Washington, DC. Available at: https://www.kmworld.com/About/What_is_Knowledge_Management. (accessed 15 February 2024).
6. Junwu, W., Yipeng, L., & Jingtao, F. Integrating Bayesian networks and ontology to improve safety knowledge management in construction behavior: A conceptual framework. *Ain Shams Engineering Journal*, 2024, vol. 15, iss. 9, article no. 102906. DOI: 10.1016/j.asej.2024.102906.
7. Moshood, D. T., Rotimi, E. F., & Rotimi, O. B. J. An Integrated Paradigm for Managing Efficient Knowledge Transfer: Towards a More Comprehensive Philosophy of Transferring Knowledge in the Construction Industry. *Construction Economics and Building*, 2022, vol. 22, iss. 3, article no. e8050. DOI: 10.5130/ajceb.v22i3.8050.
8. Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., & Duan, Z. Knowledge Graph Completion: A Review, *IEEE Access,* 2020, vol. 8, pp. 192435-192456. DOI: 10.1109/access.2020.3030076.
9. Pérez, D. S., Marbán, U. N., & Robaina, D. A. Methodology to evaluate project management and knowledge in a University. *Visión de Futuro*, 2020, vol. 24, iss. 1, pp. 114-130. DOI: 10.36995/j.visiondefuturo.2020.24.01.005.en.
10. Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., & Tahmasebi, N. Quantitative text analysis. *Nature Reviews Methods Primers*, 2024, vol. 4, iss. 1, article no. e25. DOI: 10.1038/s43586-024-00302-w.
11. Zheng, H. Vocabulary Knowledge and Acquisition. *Advances in Humanities Research*, 2024, vol. 5, iss. 1, pp. 32-36. DOI: 10.54254/2753-7080/5/2024046.
12. Gao, J., Xie, T., Li, R., Wang, K., & Zhao, L. APM: Adaptive parameter multiplexing for class incremental learning. *Expert Systems with Applications*, 2024, vol. 258, article no. 125135. DOI: 10.1016/j.eswa.2024.125135.
13. Altuhaifa, F., & Al Tuhaifa, D. Developing an Ontology Representing Fall Risk Management Domain Knowledge. *Journal of Medical Systems*, 2024, vol. 48, iss. 1, article no. e47. DOI: 10.1007/s10916-024-02062-2.
14. Luo, Q. Research on Performance Optimization Technique of Dynamic Translation System Based on Apparatus Knowledge Algorithm. *International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)*, 29–31 July 2022, Bristol, United Kingdom, IEEE, 2022, pp. 242-245 DOI: 10.1109/aiars57204.2022.00061.
15. Kundu, A., & Nguyen, U. T. Automated Fact Checking Using a Knowledge Graph-based Model. *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 19–22 February 2024, Osaka, Japan, IEEE, 2024, pp. 709-716. DOI: 10.1109/icaiic60209.2024.10463196.
16. Omran, P. G., Wang, K. & Wang, Z. An Embedding-based Approach to Rule Learning in Knowledge Graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2019, vol. 33, iss. 4, pp. 1348-1359. DOI: 10.1109/tkde.2019.2941685.
17. Tanon, P. T., Weikum, G. & Suchanek, F. YAGO 4: A Reason-able Knowledge Base. *The Semantic Web (ESWC 2020), Lecture Notes in Computer Science*, Springer, Cham, 2020, vol. 12123, pp. 583-596. DOI: 10.1007/978-3-030-49461-2_34.
18. Omran, P. G., Taylor, K., Méndez, R. S. & Armin, H. Learning SHACL shapes from knowledge graphs. *Semantic Web*, 2023, vol. 14, iss. 1, pp. 101-121. DOI: 10.3233/sw-223063.
19. Chen, G. An interpretable composite CNN and GRU for fine-grained martial arts motion modeling using big data analytics and machine learning. *Soft Computing,* 2024, vol. 28, pp. 2223-2243. DOI: 10.1007/s00500-023-09565-z.
20. Svatoš, M., Schockaert, S., Davis, J., & Kuaelka, O. Strike: Rule-driven relational learning using stratified k-entailment. *24th European Conference on Artificial Intelligence, 29 August–8 September 2020, Santiago de Compostela, Spain – Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, 2020, vol. 325, pp. 1515–1522. DOI: 10.3233/FAIA200259.
21. Oparin, G. A., Bogdanova, V. G., & Pashinin, A. A. Classification in Binary Feature Space Using Logical Dynamic Models. *44th International Convention on Information, Communication and*

Electronic Technology (MIPRO), 27 September–1 October 2021, Opatija, Croatia, IEEE, 2021, pp. 1020-1025. DOI: 10.23919/mipro52101.2021.9596697.

22. Shubin, I. Yu., Chetverykov, H. H., Snisar, S. M., & Litvin, S. H. Katehornyy analiz lohichnykh merezh dlya radiolokatsiynykh system rozpiznavannya [Categorical analysis of logical networks for radar recognition systems]. *Bionika intelektu – Bionics of Intelligence*, 2020, vol. 1, iss. 94, pp. 51-56. DOI: 10.30837/bi.2020.1(94).08. (in Ukrainian).

23. Shivappriya, S. N., Priyadarsini, M. J. P., Stateczny, A., Puttamadappa, C., & Parameshachari, B. D. Cascade Object Detection and Remote Sensing Object Detection Method Based on Trainable Activation Function. *Remote Sensing*, 2021, vol. 13, iss. 2, article no. 200. DOI: 10.3390/rs13020200.

24. Shubin I., & Kozyriev A. Method for Solving Quantifier Linear Equations for Formation of Optimal Queries to Databases. *COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems*, April 20–21, 2023, Kharkiv, Ukraine, 2023, vol. 3396, pp. 449–459. Available at: https://ceurspt.wikidata.dbis.rwth-aachen.de/Vol-3396/paper36.html. (accessed 18 July 2024).

25. Nicholls, K., Kirk, P. D. W., & Wallace, C. Bayesian clustering with uncertain data. *PLOS Computational Biology*, 2024, vol. 20, iss. 9, article no. e1012301. DOI: 10.1371/journal.pcbi.1012301.

26. Jung, P., Marra, G., & Kuželka, O. Quantified Neural Markov Logic Networks. *International Journal of Approximate Reasoning*, 2024, vol. 171, article no. 109172. DOI: 10.1016/j.ijar.2024.109172.

27. Malhotra, S., & Serafini, L. On Projectivity in Markov Logic Networks. *Machine Learning and Knowledge Discovery in Databases* (ECML PKDD 2022), Lecture Notes in Computer Science, vol 13717, Cham, Springer Nature Switzerland, 2023, pp. 223-238. DOI: 10.1007/978-3-031-26419-1_14.

Received 09.07.2024, Accepted 20.08.2024

МЕТОД ЛОГІЧНИХ МЕРЕЖ, КЕРОВАНИХ СКІНЧЕННИМИ ПРЕДИКАТАМИ ДЛЯ РОЗШИРЕНОГО АНАЛІЗУ НАВЧАЛЬНИХ ДАНИХ

З. В. Дудар, А. Д. Козирєв

Предметом дослідження є аналіз інтелектуальних даних у сфері академічної інформації. **Метою** дослідження є створення математичної моделі для аналізу навчальної інформації студентів за допомогою методу логічних мереж, керованих предикатами, що дозволяє враховувати як логічні залежності, так і ймовірнісні переходи між станами. Для досягнення поставленої мети були визначені наступні завдання: аналіз теоретичних засад методу логічних мереж і логіки предикатів, інтеграція цих підходів в єдину математичну модель, розробка підходів для її застосування в задачах класифікації академічної інформації. У дослідженні використано **методи** математичного моделювання, комплексного логічного аналізу та метод побудови логічних мереж. Отримано такі **результати**: розроблено теоретичну модель, яка об'єднує принципи логічних мереж і логіки предикатів для класифікації академічної успішності студентів; модель враховує як ймовірнісні переходи між станами, так і логічні залежності між параметрами студента; математична модель також містить логічні правила для підвищення точності класифікаційних рішень в академічному контексті. Модель було перевірено на наборі даних успішності студентів, що продемонструвало її ефективність у точному прогнозуванні навчальних результатів і підтвердило валідність інтегрованого підходу. **Висновки**. Наукова новизна отриманих результатів полягає в наступному: 1) розроблено теоретичну модель класифікації академічних даних студентів шляхом інтеграції логічних мереж і логіки предикатів, що дозволяє одночасно враховувати ймовірнісні переходи та логічні залежності між параметрами студента; 2) підхід покращує процес класифікації шляхом включення логічних правил у ймовірнісну структуру, забезпечуючи більш тонкий і точний інструмент для аналізу академічних даних; 3) ця комбінована модель пропонує новий метод для вирішення складних завдань класифікації в освітніх установах, прокладаючи шлях для подальших досліджень і практичних застосувань в інтелектуальному аналізі даних. Успішне тестування моделі на фактичних даних студентів ще більше підкреслює її потенціал як потужного інструменту для аналізу освітніх даних.

Ключові слова: інтелектуальний аналіз даних; академічні дані; логічні мережі; алгебра скінченних предикатів; логіка предикатів.

Дудар Зоя Володимирівна – канд. техн. наук, проф., завідувачка кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

Козирєв Андрій Дмитрович – асп. каф. програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

Zoia Dudar – Candidate of Technical Sciences, Associate Professor, Head of the Department of Software Engineering, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine,
e-mail: zoya.dudar@nure.ua, ORCID: 0000-0001-5728-9253, Scopus Author ID: 6506991522.

Andrii Kozyriev – PhD Student of the Department of Software Engineering, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine,
e-mail: andrii.kozyriev@nure.ua, ORCID: 0000-0001-6383-5222, Scopus Author ID: 57208898156.